

Is ChatGPT Fair for Recommendation? Evaluating Fairness in Large Language Model Recommendation

Year: 2023 | Citations: 213 | Authors: Jizhi Zhang, Keqin Bao, Yang Zhang

Abstract

The remarkable achievements of Large Language Models (LLMs) have led to the emergence of a novel recommendation paradigm — Recommendation via LLM (RecLLM). Nevertheless, it is important to note that LLMs may contain social prejudices, and therefore, the fairness of recommendations made by RecLLM requires further investigation. To avoid the potential risks of RecLLM, it is imperative to evaluate the fairness of RecLLM with respect to various sensitive attributes on the user side. Due to the differences between the RecLLM paradigm and the traditional recommendation paradigm, it is problematic to directly use the fairness benchmark of traditional recommendation. To address the dilemma, we propose a novel benchmark called Fairness of Recommendation via LLM (FaiRLLM). This benchmark comprises carefully crafted metrics and a dataset that accounts for eight sensitive attributes¹ in two recommendation scenarios: music and movies. By utilizing our FaiRLLM benchmark, we conducted an evaluation of ChatGPT and discovered that it still exhibits unfairness to some sensitive attributes when generating recommendations. Our code and dataset can be found at <https://github.com/jizhi-zhang/FaiRLLM>.